Expert-guided Machine Learning for Synthesizing $Ti0_2$ Polymorphs

Chiqun Zhang

May 2, 2017

Abstract

Background. Synthesizing complex materials can accelerate the development of technologies related to our daily life. A key bottleneck in synthesizing such materials is the large searching space of input features (such as temperature, pressure, reaction time, etc.) that needs to be explored under a limited experimental budget. Since there is not much historic data available, the success of classical machine-based algorithm is hindered. On the other hand, material science experts often have strong knowledge about what will work, and approach the problem using heuristics.

Aim. Formulate the material synthesizing as a machine learning problem combining both the accurate but expensive experiment measurement and the indirect but cheaper expert feedbacks. Build a computationally efficient algorithm to narrow search space with direct queries and pairwise comparisons. Provide guidance for new material synthesizing by predicting the desired material properties.

Data. The dataset includes two parts. One part is the experiment measurement given the combinations of the input features for the material (TiO2) synthesis. The other part of the dataset consists of the collected feedback from various material science experts in the form of pairwise comparison of the output material properties given two different input features.

Methods. In this project, we denote $f(x)$ as the unknown function mapping input x to a desired output characteristic and model the direct queries as $y = f(x) + \epsilon_D$ and the pairwise expert queries as constraints in the form of $sign(f(x) - f(x'))$. Various regression model, such as ordinary linear regression, quadratic regression and logistic regression, will be employed and compared. The second-order gradient flow algorithm will be derived and applied to find the optimized estimator.

Results. The MSE of the expert-guided machine learning algorithm is lower than the MSE of the ordinary regression methods on the same test dataset. However, the difference between the expert-guided machine learning algorithm and the ordinary regression methods is not significantly large due to the nature of the expert data, as indicated by validation dataset where the expert's estimate of the experimental outcome had very low correlation with the actual outcome and the confidence of expert being high for settings of large error.

Conclusions. By combining both the direct experiment measurement and the expert pairwise comparisons, this work demonstrates the potential capability of such a machine learning algorithm in (but not limited to) material synthesis. Also, a cross-validation method to justify the weight on the pairwise comparisons is proposed, which may be of interest in social science studies.

Keywords: Material synthesis, Machine learning, Pairwise comparison, Regression, Data analysis

DAP Committee members:

Aarti Singh (Department of Machine Learning);
Sivaraman Balakrishnan (Department of Statistics).

1 Introduction

Synthesizing complex materials can accelerate the development of technologies related to our daily life. For example, synthesizing materials with many chemical elements, engineered structures, and properties on multiple scales, from microscopic to macroscopic, can help the development of technologies from energy and environmental protection, to healthcare and communications. Nowadays, a key bottleneck in synthesizing such materials is the large space of input features that needs to be explored under a limited experimental budget. The challenge comes from the many possible combinations of input features, such as properties of the raw materials used, their mixture proportions, and processing/manufacturing settings such as temperature, pressure, and processing time, and the multiple desired characteristics of the produced material, including chemical structure, mechanical strength, and compatibility with other materials. While the classical machine learning methods can automate the multi-dimensional search, two factors hinder the success of such attempts: 1) the problem of discovering novel materials with desired properties implies there is typically not much historic data available for machine learning, and 2) due to limited budget, the number of experiments that can be conducted are small, further limiting the amount of data that can be collected for training algorithms. On the other hand, material science experts often have strong intuitions about what will work, and approach the problem using their experience and knowledge. However, these intuitions may cause unsystematic or non-diagnostic search of the many possibilities especially in high dimensions. One promising way is to develop a expert-guided machine learning algorithm that combines the experts intuitions with the experiment measurement.

2 Problem being solved

In this work, the goal is to develop an expert-guided machine learning method to control experimental parameters in synthesizing TiO_2 with desired properties. For instance, conventional processes to crystallize TiO_2 require heating in an oven at temperatures exceeding $500°C$. The goal of this project to find optimized input combinations that enable successfully synthesizing TiO_2 at low temperatures using microwaves. In this work, both the dataset from the experiments and the pairwise comparisons from the experts are employed. The exploration can be formulated as an experimental design machine learning problem where we have access to two types of noisy oracles: a fairly accurate but expensive experimental measurement, and a noisy and indirect expert feedback that is cheaper to obtain.

Formally, if $f(x)$ denotes the unknown function mapping input x to a desired output characteristic y, we model the direct and indirect queries as:

- Direct experiment queries: $y = f(x) + \epsilon_D$
- Pairwise comparison expert queries: $y = sign(f(x) - f'(x)) + \epsilon_P$

where expert feedback on uncertainty of their response will be used to learn characteristics of the noise ϵ_P. Using access to these queries, we propose to design principled and computationally efficient human-machine collaborative algorithms to demonstrate their direct utility for optimizing the input settings for low-temperature synthesis of TiO_2.

3 Background and related work

Our statistical analysis focused on a known material synthesis technique, using microwave radiation (MWR) to grow Titanium dioxide (TiO_2) thin film materials at low temperatures [Reeja-Jayan et al. (2012)]. The applications of MWR can enable growing TiO_2 on plastic-based, light-weight and flexible substrates for solar cells, light limiting diodes, sensors and photodetectors. Relevant input environment settings such as heating temperature, reaction time, and microwave power were known, but due to the lack of prior knowledge relating to MWR-assisted synthesis, the effects of the variations of the input features on the MWR result were unclear. One possible way to understand the mechanism behind the formation process is molecular simulation. However, combined with complexity of the manufacture process in MWR and the computational limitations of performing detailed simulations, molecular simulation can only deal with limited calculations in the small scale. So the hope of fast and reliable synthesis seems dim.

Another possible solution is to rely on machine learning algorithms to explore the feature space and to choose the optimized combinations of input features with desired material characteristics. In some works, the regression techniques have been demonstrated to have the capability to help choose the most promising experiment design [Nakamura et al. (2017)]. However, such method is still limited by the existing dataset and lacks the statistical support for the general case. Classical machine learning methods (for both the regression and the classification) depend on the existing data to train the model. Theoretically, the larger dataset we have for training the model, the more accurate model we get for prediction. However, in this material synthesizing problem, there is not much historical dataset in synthesizing new material and only few experiments can be conducted due to the conflict between the high expense in conducting experiment and the limited budget, by which the application of classical machine learning methods have been hindered. On the other hand, existing approaches to searching this multi-dimensional space rely on the scientific knowledge and intuition of human experts. Expert material scientists often have strong intuitions about what will work, and approach the problem using simplifying heuristics seen in discovery science [Klahr and Simon (1999, 2001); Baker and Dunbar (2000); Dunbar (1997, 1999)], and other domains (e.g., chess [Simon and Simon (1962)]). It will be much cheaper and easier to collect material science expert opinions than to conduct expensive and time consuming experiments. These intuitions can, however, lead to unsystematic or non-diagnostic search of the many possibilities [Klahr and Dunbar (1988); Schunn and Klahr (1995); Beyth-Marom and Fischhoff (1983)] especially in high dimensions, misunderstand the interpretation and communication of evidence based on prior expectations [Davis and Fischhoff (2014); Gorman (1989); Penner and Klahr (1996)], overestimate the predictability of outcomes [Fischhoff (2003)], and fail to integrate available evidence in a systematic way [Dawes et al. (1989)]. Although expert judgment and experience may be noisy, uncertain and indirect, it can be crucial and useful in developing a practical as well as reliable scheme in new complex material synthesizing. In this work, we believe a successful solution relies on leveraging both human and machine expertise and a new machine learning method combining both the direct experiment measurement and the pairwise comparisons from experts is developed and discussed.

4 Approach

In this work, we aim to build a new expert-guided machine learning algorithm that utilizes both the experiment measurement and expert pairwise comparison. The pairwise comparisons are treated as

the constraints for the optimization problem with a soft loss function. The linear regression model, quadratic regression model, logistic regression model, and support vector machine are explored and modified such that these model can accommodate the pairwise constraints. The gradient flow is used to build the equations for the model parameters, with Newton-Raphson method as the numerical solving scheme. After exploring the dataset by examining the histogram of each feature and the pairwise relation plot, the derived models are trained and tested to compare their performances. Finally, with the optimized model, we predict on the experiment grid and design profiles to help material scientist find the most promising settings for synthesizing TiO_2.

5 Method Derivation

5.1 Problem statement

Let $\hat{y}^i = f(\boldsymbol{x}^i)$, where $\hat{y}^i$ is the coverage estimate for the ith experiment, $\boldsymbol{x}^i$ is the input for the ith experiment and f is the regression function. The total experiment number is N. Then the optimization problem to find the regression function f can be written as

$$\min_f |\hat{\boldsymbol{y}} - \boldsymbol{y}|^2 \quad \text{subject to}$$
$$f(\boldsymbol{x}^i) - f(\boldsymbol{x}^j) < 0 \quad (i,j) \in \mathcal{C}$$

$\mathcal{C}$ is the comparison set. Choose the Objective function as

$$L = \sum_{i=1}^{N} \left(f(\boldsymbol{x}^i) - y^i\right)^2 + P \sum_{(i,j)\in\mathcal{C}} \left[tanh\left(\frac{f(\boldsymbol{x}^i) - f(\boldsymbol{x}^j)}{\epsilon}\right) + 1 \right],$$

where ϵ is the parameter controlling how soft the constraint loss function is and P is the penalty of the constraint loss function. In this work, ϵ will be held fixed at a value of 1 (except in Section 9) while we report results for various settings of P. Suppose f is a function with parameter $\boldsymbol{\beta}$, $(\beta_1, ..., \beta_p)$, to minimize L we need

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{0}.$$

Namely,

$$2\sum_{i=1}^{N}\left[\left(f(\boldsymbol{x}^i) - y^i\right)\frac{\partial f}{\partial \boldsymbol{\beta}}(\boldsymbol{x}^i)\right] + \frac{P}{\epsilon}\sum_{(i,j)\in\mathcal{C}}\left\{\left[\frac{1}{cosh^2\left(\frac{f(\boldsymbol{x}^i)-f(\boldsymbol{x}^j)}{\epsilon}\right)}\right]\left[\frac{\partial f}{\partial \boldsymbol{\beta}}(\boldsymbol{x}^i) - \frac{\partial f}{\partial \boldsymbol{\beta}}(\boldsymbol{x}^j)\right]\right\} = 0 \quad (1)$$

5.2 Linear Regression

If assume $f(\boldsymbol{x}^i) = \boldsymbol{x}^i \cdot \boldsymbol{\beta}$, or $\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\beta}$, then Eqn 1 becomes

$$2\left(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T\boldsymbol{y}\right) + \frac{P}{\epsilon}\sum_{(i,j)\in\mathcal{C}}\left\{\frac{\boldsymbol{x}^i - \boldsymbol{x}^j}{cosh^2\left(\frac{(\boldsymbol{x}^i-\boldsymbol{x}^j)\cdot\boldsymbol{\beta}}{\epsilon}\right)}\right\} = 0 \quad (2)$$

Left hand side of Eqn 2 is denoted as Residue, $\boldsymbol{r}$. The Hessian $\boldsymbol{H}$ for this problem will be

$$\boldsymbol{H} = 2\boldsymbol{X}^T\boldsymbol{X} - \frac{2P}{\epsilon^2}\sum_{(i,j)\in\mathcal{C}} \frac{sinh\left(\frac{(\boldsymbol{x}^i-\boldsymbol{x}^j)\cdot\boldsymbol{\beta}}{\epsilon}\right)}{cosh^3\left(\frac{(\boldsymbol{x}^i-\boldsymbol{x}^j)\cdot\boldsymbol{\beta}}{\epsilon}\right)}(\boldsymbol{x}^i-\boldsymbol{x}^j)\otimes(\boldsymbol{x}^i-\boldsymbol{x}^j) \tag{3}$$

For gradient flow, the update rule should be

$$\boldsymbol{\beta}^{t+1} = \boldsymbol{\beta}^t - \gamma\boldsymbol{H}^{-1}\boldsymbol{r}^t,$$

where γ is the step size.

In addition to the ordinary linear regression, the quadratic regression and the logistic regression are also be embedded into this scheme. For the quadratic regression, the feature space is expanded to the second order of every features. For the logistic regression, we treat the regression function as one variation of the ordinary linear regression. The inverse of the logistic function, $\boldsymbol{g}$, is defined as

$$g(\boldsymbol{y}) = log(\frac{\boldsymbol{y}}{1+\boldsymbol{y}}).$$

For the logistic regression, we assume $g(\boldsymbol{y}) = \boldsymbol{x}^i\cdot\boldsymbol{\beta}$ and approximately solve the problem with minimizing the least square error.

5.3 Classification

In classification problem. the loss function is given as

$$L = \frac{1}{2}||\boldsymbol{\beta}||^2 + C\xi^T\mathbf{1}_n - \sum_{i=1}^n \alpha_i y_i < \boldsymbol{x}^i, \boldsymbol{\beta} > + \boldsymbol{\alpha}^T\mathbf{1}_n - \xi^T(\boldsymbol{\alpha}+\eta) + P\sum_{(i,j)\in\mathcal{C}}\left[tanh\left(\frac{f(\boldsymbol{x}^i)-f(\boldsymbol{x}^j)}{\epsilon}\right)+1\right],$$

where $\xi = max(1 - y_i < \boldsymbol{x}^i, \boldsymbol{\beta} >, 0)$, $C = \frac{1}{\lambda}$, and $\boldsymbol{\alpha}$ and η are two Lagrange parameters. From the dual form and taking the derivative with respect to $\boldsymbol{\beta}$, ξ and α, we have

$$\boldsymbol{\beta} - \sum_{i=1}^n \alpha_i y_i \boldsymbol{x}^i + \frac{P}{\epsilon}\sum_{(i,j)\in\mathcal{C}}\left\{\frac{\boldsymbol{x}^i-\boldsymbol{x}^j}{cosh^2\left(\frac{(\boldsymbol{x}^i-\boldsymbol{x}^j)\cdot\boldsymbol{\beta}}{\epsilon}\right)}\right\} = 0$$

$$\alpha = \text{argmax}\boldsymbol{\alpha}^T\mathbf{1}_n - \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Y}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{\alpha}$$

Therefore, we first get the estimate of $\boldsymbol{\alpha}$ and then use Newton-Raphson method to get the $\boldsymbol{\beta}$.

6 Data Description

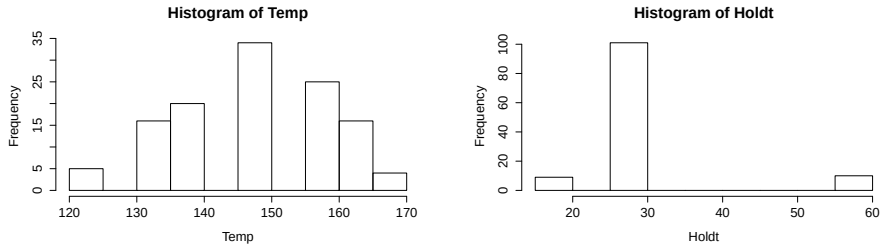
6.1 General Description

The dataset includes two parts. One part is the measurement of the input features for the material (TiO_2) synthesis experiments and the output material property. Now, we have totally 124 experiment observations, whose feature variables are Temperature, Ramp Time, Hold Time and TEG Ratio. The label or output of this dataset is the coverage percentage of TiO_2 observed. By adjusting the combination of these four environment settings, different TiO_2 coverages will be

observed. The other part of the dataset consists of the collected feedback from several material science experts in the form of the pairwise comparison of the output material properties given two different input features. There are 142 pairs of comparisons from experts corresponding different environment settings, 102 entries of which are collected from three experts and have tracked each expert's confidence for their answers as well as direct coverage estimated by the expert.

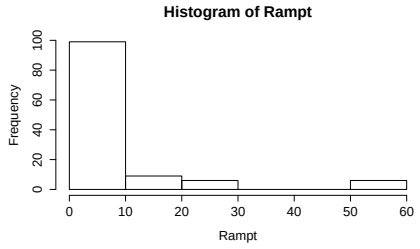
6.2 Exploratory Data Analysis

Figure 1 is the histograms of the predictor and label variables. The distribution of $RampTime$ is left skewed and needs to be transformed. In this work, we apply the log transformation on $RampTime$. And the histogram of the $log(RampTime)$ is shown as Figure 1(f), which is more normally distributed.

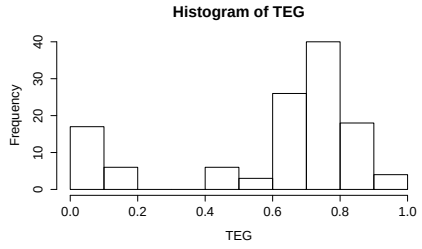


(a) Histogram of Temp

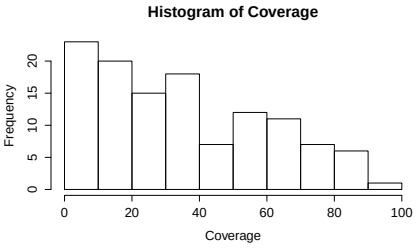
(b) Histogram of HoldTime



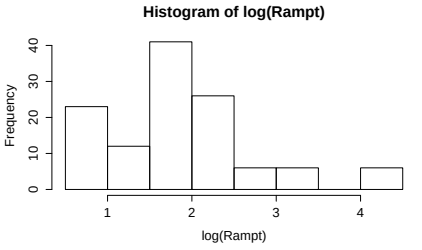
(c) Histogram of RampTime



(d) Histogram of TEG



(e) Histogram of Coverage



(f) Histogram of Log(RampTime)

Figure 1: Histogram for all variables.

To get a general understanding of the correlation within the dataset, we generate the pairwise plot. Figure 2 is the pairwise plot between $Coverage$ and other predictors. Due to the complexity of this dataset, there is no obvious pattern between the $Coverage$ and other features. However, the

pairwise plot between the $Temp$ and $Coverage$ shows that $Temp$ likely has a quadratic relationship with $Coverage$. We will show the results for both the linear regression and the quadratic regression in Section 7.

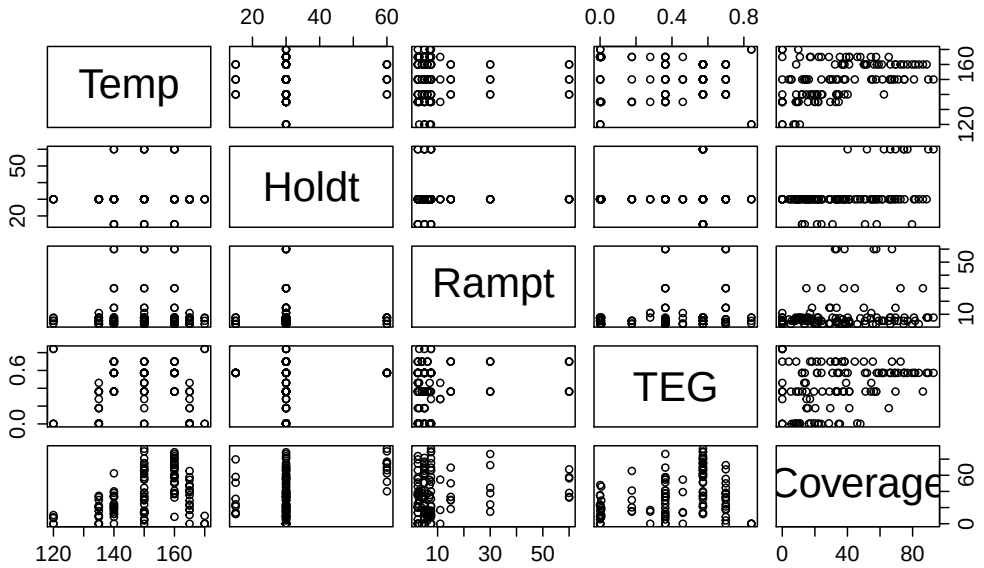


Figure 2: Pairwise plot between Coverage and other predictors

7 Experiments

In this part, we apply the regression models in Section 5 to predict the material coverage. The results from the model without pairwise comparisons and the model with the pairwise comparisons are compared. In all calculations, the weight value P is chose by the cross-validation: at every P, the experiment observations data is randomly divided into the training part (80% of the dataset) and the validation part (20% of the dataset) for 100 times to calculate the mean of mean square error and the variance of the mean square error. P with the least mean of MSE is chosen as the optimized P and is used in training the model. In addition, based on the variance of the MSE at every P, the 95% confidence interval of the mean of MSE at every P is calculated with the Gaussian distribution assumption. In Section 7.1, we use the linear predictors, $Y \sim Temp + HoldT + RampT + TEG$ ($Temp$ stands for $Temperature$, $HoldT$ for $Hold\,Time$, $RampT$ for $Ramp\,Time$ and TEG for $TEG\,ratio$). In Section 7.2, we discuss the results of quadratic predictor model $Y \sim Temp + HoldT + RampT + TEG + Temp^2 + HoldT^2 + RampT^2 + TEG^2$ (the reason we discard the interactions between features is all features in this problem are prescribed independently in experiments). In Section 7.3, the variation of the logistic regression with quadratic predictors are explored and discussed. All results presented in this work are based on the whole direct experiment measurement and all pairwise comparisons, unless otherwise specified. ϵ is assumed as 1 in all calculations except in Section 9. Note that ideally the optimized ϵ should also be found through the cross-validation process. In this section, the ordinary model without pairwise comparisons is

denoted as the ordinary model and the model proposed with the pairwise comparisons is called the expert-guided model.

7.1 Linear predictors

7.1.1 Regression error

We start from the linear predictor. Figure 3 shows the cross-validated mean of MSE at different P values, indicating that 0.96 is the optimized P value. The red line in Figure 3 is the mean of MSE from the ordinary model, the red dashed lines above and below are the upper bound and lower bound corresponding to a 95% confidence interval. The blue stars are the mean of MSE from the expert-guided model at different P values, and the blue circles above and below are the upper and lower bound corresponding to a 95% confidence interval. It shows that the expert-guided model has the slightly lower mean of MSE at the optimized P. At the optimized P, the mean of MSE is 260.9 for the ordinary model and 260.7 for expert-guided model.

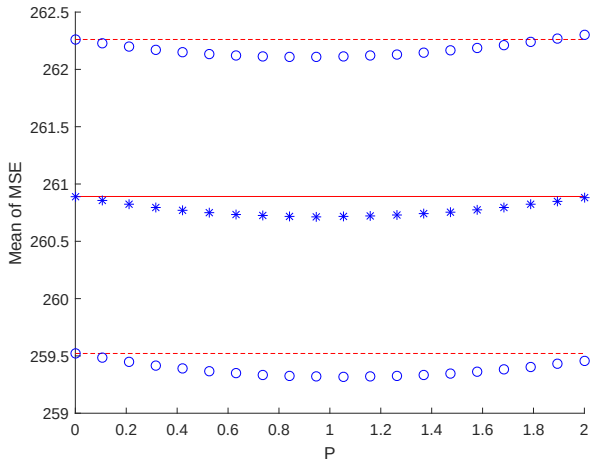


Figure 3: Mean of MSE at different P for the linear predictor model. Both the mean and the 95% confidence interval are presented.

7.1.2 Classification error

In some cases, material experts are also interested in if a particular input will lead to new material covering area big enough rather than the actual synthesizing coverage, which transfers the regression problem into a classification problem. In the classification problem, we adopt the scheme proposed in Section 5.3. We treat the coverage greater than 50% as the $Class$ 1 and the coverage less than 50% as $Class$ 2. With the optimized P through cross validation, there is no difference between the ordinary model and the expert-guided model, both accuracies are 42%.

7.2 Quadratic predictors

In addition to exploring more accurate and efficient model, the motivation to utilize the quadratic predictors is that linear predictors will always predict the maximum coverage at the extreme val-

ues. However, the goal of this work is to find the optimized combination of features with highest material coverage. Therefore, in this part, we extend the predictors from linear to quadratic. Cross-validation is applied again to find the optimized P as stated at the beginning of Section 7. Figure 4 shows the mean of MSE at different P values from quadratic model. The red lines and dashed lines, as well as the blue stars and circles stand for the same meanings as in Section 7.1. It is clear that the expert-guided model still has the lower mean of MSE at the optimized P. The optimized P is 0.54. At the optimized P, the mean of MSE is 261 for the ordinary model and 260.8 for the expert-guided model. In addition, if we use these two regression models as the classifier, the classification accuracies are the same, 47%.

Comparing the models with the linear predictor and with the quadratic predictor, we find that these two models give similar performance on the same dataset. However, due to the motivation argued above, we will apply the quadratic predictor in the prediction.

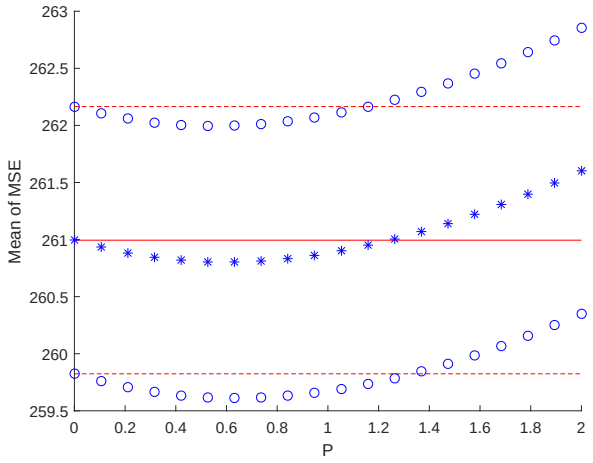


Figure 4: Mean of MSE at different P from quadratic predictor model. Both the mean and the 95% confidence interval are presented.

7.3 Logistic regression on the dataset

In the linear regression and quadratic regression, the predicted coverage is not guaranteed to lie between 0 to 100, which is not physically reasonable. To solve this problem, we try to convert the problem into a logistic regression by interpreting the coverage as the possibility of the desired material covering the whole sample. Namely, we convert the coverage percentage into the range between 0 and 1 and apply the logistic regression so that the predicted coverage will not go beyond 100%. As discussed in Section 5, we find the optimized estimator by minimizing the least square error rather than maximizing the maximum likelihood. With the logistic model, we try to predict coverage on grid and design profiles. To avoid getting the maximum at the extreme values, we apply the quadratic predictors.

7.3.1 Choose the optimized P

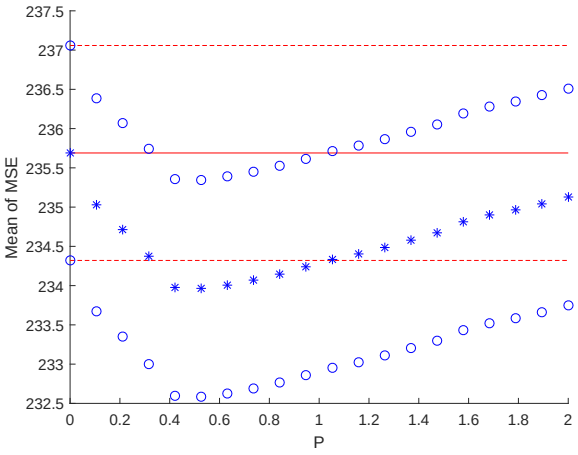


Figure 5: Cross-validated mean of MSE at different P from logistic model. Both the cross-validated mean and the 95% confidence interval are presented.

Figure 5 is the plot between different P values and the mean of MSE. Both the cross-validated mean of MSE and the 95% confidence interval are presented. It shows that the optimized P is 0.42. Also, the mean of MSE from the logistic model is lower than the linear regression model and the quadratic regression model, indicating the logistic model is more accurate in predicting the coverage by constraining the prediction between 0 and 100.

7.3.2 Predict on selected experiment grid profile

In this part, the logistic regression model is employed to predict on selected conditions from the grid profile and compare with the experiment measurement. The grid profile is generated by subsampling the range of each predictor. The range for each feature is $0-200$ for $Temperature$, $0-120$ for $RampTime$, $0-60$ for $HoldTime$, and $0-1$ for $TEGratio$. Ten arbitrary environment conditions that do not appear in the experiment measurement dataset are selected as the prediction input. Also, the experiment measurement of these ten conditions are then acquired by the material science collaborators and are not included in the training dataset. The selected ten conditions are as following

Temp	HoldTime	RampTime	TEG
118	57	120	0.8
131	57	114	0.8
145	32	114	0.8
158	6	107	0.8
158	57	107	0.8
158	6	114	0.8
158	19	120	0.8
158	44	120	0.8
158	57	95	0.8
198	6	120	0.8

Figure 6 shows the relation between the predictions and experiment measurement for these conditions from the grid. It shows that most of the conditions with high predicted coverage also have relative high experiment measurement. But the prediction accuracy in terms of the coverage percent is not high.

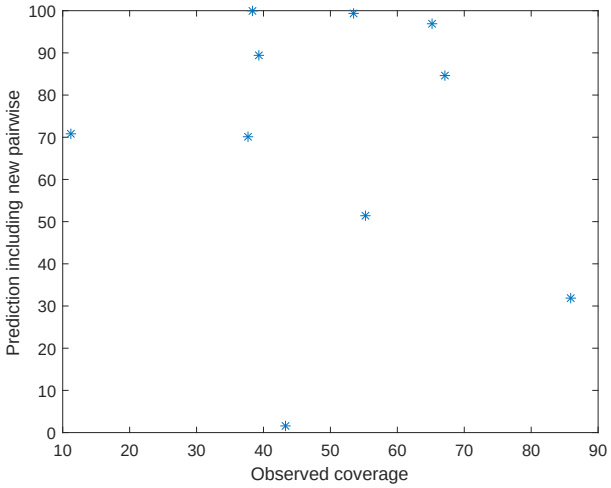


Figure 6: The relation between the predictions and the observed coverages.

7.3.3 Predict on experiment grid profile and experiment design profile

With the logistic regression expert-guided model, we predict the coverage on all of the grid profile generated in Section 7.3.2 and the experiment design profile that are set of experiments the collaborator was planning to do next. For each profile, the top ten conditions with the highest predicted coverage are selected and presented in this section.

For the generated grid profile, the top ten conditions with the maximum coverage are

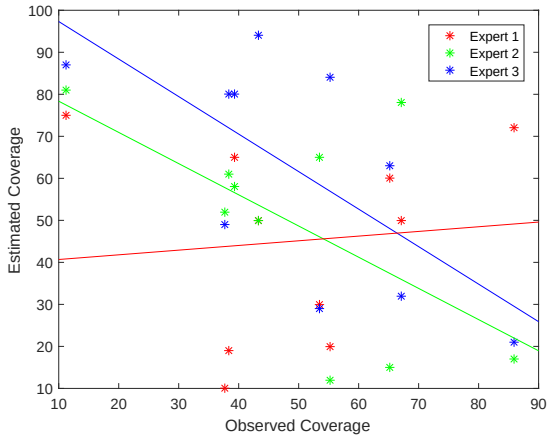
Temp	HoldTime	RampTime	TEG
158	120	40	0.8
163	120	40	0.8
158	120	53	0.8
163	120	53	0.8
153	120	40	0.8
168	120	40	0.8
153	120	53	0.8
168	120	53	0.8
158	120	27	0.8
163	120	27	0.8

These are close to the settings which led to largest coverage as obtained in an independent experiment set (previous table) - $Temp$ at 158, $HoldTime$ at 120, $RampTime$ at 19 and TEG at 0.8. For the experiment design profile, the top ten conditions with the maximum coverage are

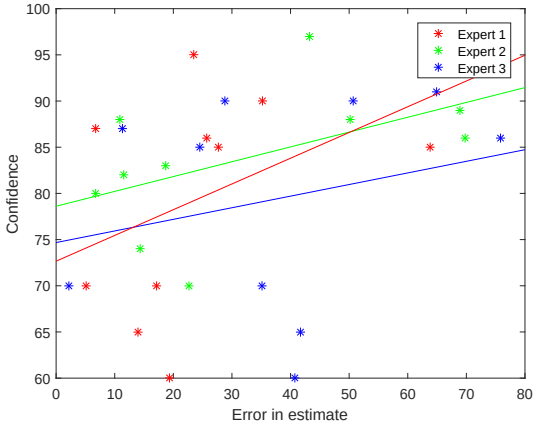
Temp	HoldTime	RampTime	TEG
160	115	30	0.8
160	115	60	0.8
150	115	30	0.8
150	115	60	0.8
160	115	15	0.8
150	115	15	0.8
135	115	30	0.8
185	115	30	0.8
135	115	60	0.8
185	115	60	0.8

Since the optimized temperatures in the predictions on the grid and design profiles are localized, the predictions indicate the model proposed in this work is reasonable.

8 Diagnosis of the pairwise comparisons



(a) Correlation between the observed coverage and experts' estimate.



(b) Plot between the estimation error and experts' confidence.

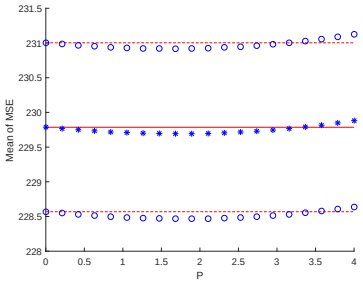
Figure 7: Relation between the experiment measurement and expert estimate and the confidence compared with estimation errors.

From the results shown in Section 7, it shows that the difference of the mean of MSE between the ordinary model and the expert-guided model is not very large. Therefore, we also explore the pairwise comparisons to learn the performance of the expert estimate. We picked out the data entries that exist both in the experiment measurement and in the pairwise comparisons. Figure 7(a) shows the relation between the available experiment measurement and their corresponding direct coverage estimate from three experts. The lines in Figure 7 are the least square fit lines corresponding to each expert. The correlation coefficients between the experiment measurement and the estimate for each expert are 0.09, -0.59, and -0.67 respectively. In Section 9, we will explore and discuss the regression model with individual expert comparisons. Figure 7(b) shows the

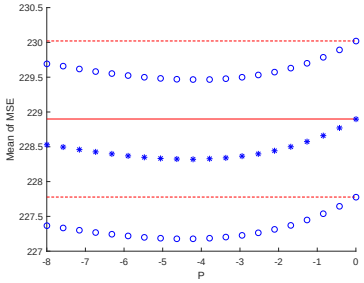
confidence (ranging from 0 to 100) provided by experts and their errors in estimate. Interestingly, the larger error one expert makes, the more confident that expert is, which may be of interest in social science or human behaviors studies.

9 Prediction with pairwise comparisons from Expert 1,2&3

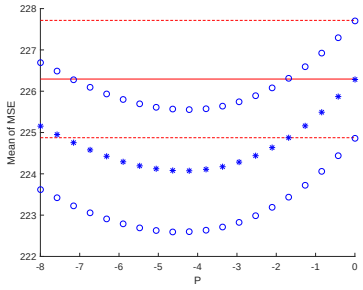
In this part, we explore the model with individual pairwise comparisons. After examining the results with each expert comparisons, we find that the assumed ϵ being 1 is no longer robust and reasonable. Therefore, we assume a more soft constraint loss and set ϵ being 10 in this section. Recall that Expert 1 is the only expert with the positive correlation between his/her estimate and the experiment measurement, as discussed in Section 6.2. First, we use Expert 1 pairwise comparisons only and re-train the logistic regression model. Figure 8(a) shows the cross-validated mean of MSE with different P values with Expert 1 pairwise comparisons. Similarly, Figure 8(b) and Figure 8(c) show the mean of MSE at different P values from the logistic model with Expert 2 pairwise comparisons and Expert 3 pairwise comparisons respectively.



(a) Mean of MSE at different P with Expert 1 comparisons.



(b) Mean of MSE at different P with Expert 2 comparisons.



(c) Mean of MSE at different P with Expert 3 comparisons.

Figure 8: Mean of MSE at different P values of logistic regression settings with Expert 1, Expert 2 and 3 comparisons.

From Figure 8, it shows that $P_3 < P_2 < 0 < P_1$ where P_i denotes the optimized P for Expert i pairwise comparisons. Also, the mean of MSE from the model with Expert 3 comparisons is the lowest and the mean of MSE from the model with Expert 1 comparisons is the highest. The results generated from the proposed model match with the correlation coefficient presented in Section 8.

In addition, we generate the top 10 promising conditions from the experiment design profile from the logistic model with only Expert 1 comparisons. The results are shown as the following.

Temp	HoldTime	RampTime	TEG
160	115	30	0.8
160	115	60	0.8
150	115	30	0.8
150	115	60	0.8
160	115	15	0.8
150	115	15	0.8
185	115	30	0.8
135	115	30	0.8
185	115	60	0.8
135	115	60	0.8

10 Discussion and Conclusion

An expert-guided machine learning algorithm is proposed in this work, taking the advantages of both the direct experiment measurement and the experts pairwise comparisons. Linear regression, quadratic regression, logistic regression and SVM classification methods are embedded into this scheme and their second order gradient flow numerical schemes are given respectively. Material synthesizing problem is explored and discussed with the expert-guided model including exploratory data analysis, parameter optimization with cross-validation and prediction on design and grid profile.

From the results discussed in this work, we show that the pairwise comparisons are helpful in developing the prediction model with lower error by choosing the proper parameters. It also shows that the prediction accuracy in terms of the coverage percent is not very well due to the data complexity in this project. The correlation coefficients between the experts estimate and the experiment measurement indicate the high bias and variance in the experts pairwise comparisons, which emphasizes the significance of the combination with the direct measurement. Although the theoretical mechanism behind is not clear now, the scheme proposed in this work is general and capable to various applications, especially the cases with few dataset available.

11 Possible Future Work

The future work is to embed more regression or classification methods into the proposed scheme and to apply on more various applications. The theoretical studies to determine when the pairwise comparisons will help positively and to derive analytical confidence intervals are also of great interest.

12 Lessons Learned

First, the experience in handling dataset from real experiments with solid and meaningful motivation. One of the critical intention of the data analysis project to let students apply the machine learning techniques to real-world dataset. During this project, I get my hand on the direct dataset

from material science labs and the pairwise comparison data from material science expert surveys, both of which are with noise, bias and uncertainties. From this project, I have learned the basic work flow to deal with such datasets, from pre-process and cleaning up to interpreting the physical meanings of the output. The experience acquired from this project will be helpful in my future career as a data scientist.

Second, a deeper understanding on the classical machine learning algorithms and their limitations and advantages. In addition to processing the complicated dataset, the novelty of this work is to develop a new machine learning algorithm that combines both the direct dataset and the pairwise comparisons. During the derivation of the new algorithm, I reviewed the classical techniques in machine learning and have acquired a deeper understanding on their limitations and applications. Also, the experience in testing the algorithm and adjusting the model parameter is also valuable and helpful.

Last, a stronger ability in communication and collaboration with other scholars. This work is collaborated with professors and scholars from Department of Machine Learning, Material Science Department and Department of Engineering and Public Policy. From discussing and sharing ideas with people from different backgrounds, I strengthen my communication skills and learn how a collaborated project is conducted.

13 Acknowledgements

This research is supported in part by an INCUBATE Seed Grant, Carnegie Mellon University - Expert-guided machine learning for nanomaterial discovery. I would like to thank Professor Aarti Singh and Professor Sivaraman Balakrishnan for the support and advice in this project. I really appreciate Professor Reeja Jayan and her group for conducting the experiments and collecting the datasets, and Professor Alex Davis for designing the survey and generating the pairwise comparisons. Also, I want to thank Elijah Peterson and Yining Wang for the technical discussions. The continuous support and help from Diane Stidle and Dorothy Holland-Minkley are highly appreciated.

References

Baker, L. M. and Dunbar, K. (2000). Experimental design heuristics for scientific discovery: The use of baseline and known standard controls. *International Journal of Human-Computer Studies*, 53(3):335–349.

Beyth-Marom, R. and Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of personality and social psychology*, 45(6):1185.

Davis, A. L. and Fischhoff, B. (2014). Communicating uncertain experimental evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1):261.

Dawes, R. M., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899):1668–1674.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. *Creative thought: An investigation of conceptual structures and processes*, 4.

Dunbar, K. (1999). How scientists build models in vivo science as a window on the scientific mind. In *Model-based reasoning in scientific discovery*, pages 85–99. Springer.

Fischhoff, B. (2003). Hindsight≠ foresight: the effect of outcome knowledge on judgment under uncertainty. *Quality and Safety in Health Care*, 12(4):304–311.

Gorman, M. E. (1989). Error, falsification and scientific inference: An experimental investigation. *The Quarterly Journal of Experimental Psychology*, 41(2):385–412.

Klahr, D. and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science*, 12(1):1–48.

Klahr, D. and Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125(5):524.

Klahr, D. and Simon, H. A. (2001). What have psychologists (and others) discovered about the process of scientific discovery? *Current Directions in Psychological Science*, 10(3):75–79.

Nakamura, N., Seepaul, J., Kadane, J. B., and Reeja-Jayan, B. (2017). Design for low-temperature microwave-assisted crystallization of ceramic thin films. *Applied Stochastic Models in Business and Industry*.

Penner, D. E. and Klahr, D. (1996). When to trust the data: Further investigations of system error in a scientific reasoning task. *Memory & cognition*, 24(5):655–668.

Reeja-Jayan, B., Harrison, K. L., Yang, K., Wang, C.-L., Yilmaz, A., and Manthiram, A. (2012). Microwave-assisted low-temperature growth of thin films in solution. *Scientific reports*, 2.

Schunn, C. D. and Klahr, D. (1995). A 4-space model of scientific discovery. In *Proceedings of the 17th annual conference of the cognitive science society*, pages 106–111.

Simon, H. A. and Simon, P. A. (1962). Trial and error search in solving difficult problems: Evidence from the game of chess. *Behavioral Science*, 7(4):425.